JOURNAL OF CLINICAL MICROBIOLOGY, July 1997, p. 1751–1756 Vol. 35, No. 7
0095-1137/97/$04.00+0

Interlaboratory Agreement among Results of Human Papillomavirus Type 16 Enzyme-Linked Immunosorbent Assays

HOWARD D. STRICKLER,[1]* ALLAN HILDESHEIM,[1] RAPHAEL P. VISCIDI,[2] KEERTI V. SHAH,[2] BRAD GOEBEL,[3] JAMES DRUMMOND,[3] DAVID WATERS,[3] YEPING SUN,[3] NANCY L. HUBBERT,[3] SHOLOM WACHOLDER,[1] LOUISE A. BRINTON,[1] CHENG-LONG HAN,[1] PHILIP C. NASCA,[4] ROBERTA MCCLIMENS,[5] KAREN TURK,[6] VIOLET DEVAIRAKKAM,[6] SUSAN LEITMAN,[7] CYNTHIA MARTIN,[7] AND JOHN T. SCHILLER[1]

National Cancer Institute[1] and Department of Transfusion Medicine,[7] National Institutes of Health, Bethesda, The Johns Hopkins Medical Institutions, Baltimore,[2] SAIC, Frederick,[3] and Information Management Services Inc., Rockville,[5] Maryland; School of Public Health and Health Sciences, University of Massachussets, Amherst, Massachusetts[4]; and Research Triangle Institute, Washington, DC[6]

Received 15 November 1996/Returned for modification 27 January 1997/Accepted 7 April 1997

Serological assays for measuring antibodies to human papillomavirus type 16 (HPV-16) virus-like particles (VLPs) have become important epidemiologic tools in recent years. However, the interlaboratory replicability of these assays has not been assessed. In this investigation, three laboratories tested a panel of specimens obtained from two different groups: 265 subjects in a vulvar cancer case-control study and 107 healthy volunteer blood donors. Each laboratory used an enzyme-linked immunosorbent assay (ELISA), but no attempt was made to standardize assay procedures among the three laboratories. The data showed good day-to-day intralaboratory replicability in laboratory 1 (correlation coefficient, ≥0.88) and good intra-assay variability in laboratory 3 (correlation coefficient, ≥0.93). Interlaboratory correlations, likewise, ranged between 0.61 and 0.80 in both case-control study subjects and healthy blood donors, indicating that ELISA optical density (OD) values between laboratories were linearly related regardless of the population. Kappa coefficients (κ), based on each laboratory's categorical interpretation of its results (as positive or negative), showed good agreement (κ, >0.6) in case-control study subjects and moderate agreement (κ, ≥0.4) in blood donors, a population that had few strongly positive sera. When OD values near seropositive cutoffs were treated as indeterminates, there was little discordance between laboratories in either population. The data suggest that each laboratory measured the same humoral immune response and that their HPV-16 VLP ELISAs performed similarly (Pearson correlations). Interlaboratory differences, however, probably due to reagents and procedures, were considerably greater than intralaboratory day-to-day variability. Interlaboratory agreement in determining seropositivity (κ) could be improved by sharing positive and negative serum controls and by treating marginal results as indeterminate. As part of continuing cooperation to improve interlaboratory agreement, we are preparing bulk serum control specimens to be shared and made available to interested researchers.

Since their introduction in 1992, assays to measure antibody responses to human papillomavirus (HPV) virus-like particles (VLPs) have become increasingly important research tools (4, 9, 11, 12, 15). Sexually transmitted HPV infections are widely accepted as causing cervical cancer and its precursors, squamous intraepithelial lesions (7). However, the HPV types that are a high risk for the development of cancer cannot be grown preparatively in culture. The production of HPV VLPs by recombinant DNA technology allowed researchers for the first time to develop enzyme-linked immunosorbent assays (ELISAs) using conformationally intact viral capsid proteins from high-risk HPV types (11), the most important of which is HPV type 16 (HPV-16) (7).

Detection of HPV-16 VLP immunoglobulin G (IgG) by ELISA is strongly associated with detection of cervical neoplasia and cervical HPV-16 DNA (2, 11, 12, 15, 19). Likewise, HPV-16 VLP antibodies are elevated in patients with cancers of the anus, vulva, vagina, and penis, in accord with mounting DNA hybridization evidence that these tumors are also HPV related (5–7, 14). To help determine which additional tumors might be HPV related, researchers have recently conducted large serological surveys of cancer patients (14). Similarly, serological surveys are being used in population-based studies to assess whether regional (ecological) differences in exposure to HPV-16 might help explain international differences in anogenital cancer incidence (12).

HPV-16 VLP serology is useful, therefore, in conducting studies in which DNA detection methods are not practical because of unavailability of tissue specimens or because of cost (e.g., population surveys) and as an independent measure of virus exposure (e.g., to verify purported cancer associations with detection of HPV-16 DNA). There is also speculation that HPV-16 VLP serology may have clinical utility, for example, as an adjunct to the Pap smear in cervical cancer screening programs (13). However, HPV-16 VLP serological assays are still at a developmental stage. No commercial test is available. More fundamentally, no studies have directly compared results in different laboratories to measure interlaboratory agreement or to assess whether the same humoral immune response is being measured in each laboratory.

To measure interlaboratory agreement in the HPV-16 VLP ELISA, we compared the results obtained by three laboratories with a set of identical serum specimens obtained from two heterogeneous groups of subjects, including patients with vulvar cancer, a tumor that frequently contains HPV-16. This

* Corresponding author. Mailing address: Viral Epidemiology Branch, NCI, NIH, EPN Room 434, Bethesda, MD 20892. Phone: (301) 496-8115. Fax: (301) 402-0817.

report describes each laboratory's ELISA methods, several measures of the level of interlaboratory agreement, and plans for further enhancing reproducibility between laboratories.

MATERIALS AND METHODS

Subjects. (i) Vulvar cancer cases and controls. Case subjects were 142 women patients with histologically confirmed vulvar cancer or vulvar intraepithelial neoplasia III, diagnosed between 1985 and 1987, enrolled in a multicenter investigation based in hospitals throughout upstate New York and Chicago. Controls were 126 community-based subjects who were matched to case subjects by sex, age (±5 years), race, and telephone exchange or ZIP code and enumerated by using random-digit dialing techniques (controls under the age of 65 years) or by using information provided by the Health Care Financing Administration (controls older than age 65) (16). A previous report showed that HPV-16 VLP antibodies were more prevalent in case than in control subjects (6).

(ii) Blood donors. Sequential healthy male and female volunteer blood donors (n = 107) presenting to the National Institutes of Health and enrolled in a program to collect blood specimens for research purposes provided samples for testing in the current investigation.

Serology. Three independent laboratories agreed to test serum specimens for IgG to HPV-16 VLPs by using their in-house ELISAs. No special instructions were given to the laboratories regarding the handling or testing of these specimens. In addition, each laboratory was masked to patient characteristics and the other laboratories' results. The HPV-16 VLPs used in all three laboratories were prepared from Sf9 insect cells infected with a recombinant baculovirus expressing the L1 and L2 proteins of HPV-16 (strain 114/K) and purified by the method of Kirnbauer et al. (10). Laboratories 1 and 3 each produced their own VLPs, whereas laboratory 2 used VLPs produced by laboratory 1.

Several important characteristics of each laboratory's ELISA methods are summarized in Table 1. In particular, note that in laboratory 1 serum specimens were tested in single wells for blood donors. However, sera from the case-control study subjects were tested in triplicate, once each on 3 separate days. Thus, in laboratory 1, the replicate results for case-control study subjects are measurements of true intralaboratory (day-to-day) variability rather than intra-assay variability. The mean optical density (OD) of the triplicates was used for interlaboratory comparisons. In laboratory 3, specimens were tested in duplicate on the same plate. Thus, in laboratory 3, the replicate values reflect intra-assay variability. The ELISA value analyzed was the mean OD of the duplicates for a given specimen measured on adequately tested plates. Specimens in laboratory 2 were tested once each. Detailed procedures for laboratories 1 and 3 have been previously published (11, 17). Therefore, only the methods for laboratory 2 are described in full here.

Laboratory 2 ELISA. HPV-16 VLPs were diluted to 1 μg/ml in phosphate-buffered saline (PBS). Wells in a 96-well polystyrene plate (Nunc, Naperville, Ill.) were then coated with 0.1 ml of the VLP solution, incubated overnight at 4°C, and washed three times with wash solution (0.05% Tween 20 and 0.01% Triton X-100 [Sigma, St. Louis, Mo.] in PBS) in an automatic plate washer (Microwash 2; Skatron, Lier, Norway). The plate was tapped dry on paper towels, and 300 μl of a blocking solution composed of 5% bovine serum albumin (Sigma) in PBS was added to each well. The plate was incubated at 37°C for at least 3 h and then washed three times as before. Human serum samples, diluted 1:20 in blocking solution, were next added at 100 μl per well, incubated on the plate for 1 h at 37°C under a plate sealer, and then washed five times. To detect bound antibody, goat anti-human IgG conjugated with alkaline phosphatase (Boehringer Mannheim Corp., Indianapolis, Ind.) and diluted 1:3,000 in blocking solution was added to each well (100 μl), incubated under a plate sealer for 30 min at 37°C, and then washed. Substrate (1 mg of 1-*para*-nitrophenylphosphate per ml in diethanolamine buffer [pH 9.8; Sigma]) was added at 100 μl per well and incubated while covered for 30 min at 37°C. The enzyme reaction was subsequently stopped by addition of 50 μl of 3 N NaOH per well. The plate was read at 405 nm in an automated ELISA plate reader (EL312e; BIO-TEK; Winooski, Vt.). Each specimen was tested in a single well.

Statistical methods. We used Spearman and Pearson correlation coefficients (ρ), as well as kappa coefficients (κ), to measure interlaboratory agreement, as each method has different limitations. In general, Pearson correlation coefficients measure the linearity of the relationship between two sets of values. However, this method is sensitive to the distribution of the data (8). To normalize OD distributions and minimize the effects of any outliers, we log transformed the OD values in Pearson correlation analyses. Spearman correlations depend only on the order of the observations, not their actual values, and therefore do not consider whether the relationship is linear. Both Spearman's and Pearson's correlations are affected by data spread. The two methods gave almost the same results with our data sets, and we present Pearson coefficients because they are directly interpretable as simple regression estimates, as presented in the figures. Kappa coefficients generally measure the degree of agreement beyond chance between two or more categorical ratings of a given factor, but kappas are affected by the choice of seropositive cutoff and underestimate agreement in skewed distributions (18).

Categorical (positive and negative) results in the assays were determined by using cutoff values pre-established in laboratories 1 and 3. In laboratory 2,

TABLE 1. Brief comparison of laboratory methods

Laboratory	HPV-16 VLP concn (μg/ml) in PBS (vol [μl])	Microtiter plate	Blocking agent	Dilution of serum (vol [μl])	Serum controls	Replicates
1	10 (50)[a]	Immulon II[b]	0.5% nonfat dry milk + 0.1% calf serum in PBS	Serum 1:4 in 0.5% dry milk–PBS (50)	Positive + negative	Triplicate[c] testing of case-control specimens
2	1 (100)[a]	Nunc-Polysorp[d]	5% BSA in PBS	Serum 1:20 in blocking solution (100)	None	Single tests
3	5 (50)	Nunc-Polysorp	None	Serum 1:10 in 0.5% dry milk–PBS (100)	Positive + negative	Duplicate[c] testing of all specimens

[a] VLPs were produced by laboratory 1.
[b] Dynatech, Chantilly, Va.
[c] Testing of case-control specimens in laboratory 1 was conducted in triplicate, once each on 3 separate days. Thus, in laboratory 1, the replicate results for case-control subjects measured true intralaboratory (day-to-day) variability, rather than intra-assay variability. In laboratory 3, specimens were tested in duplicate on the same plate. Thus, in laboratory 3, the replicate values reflect intra-assay variability.
[d] Nunc, Naperville, Ill.

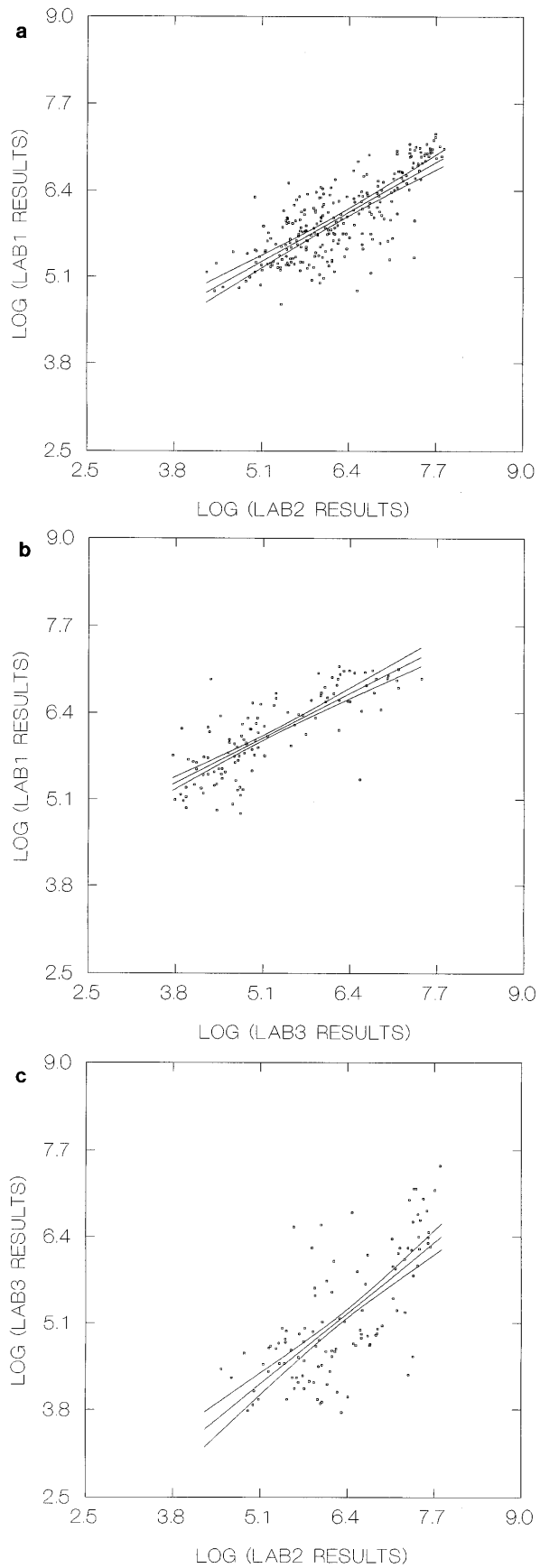


FIG. 1. Comparison of HPV-16 VLP ELISA OD values for vulvar cancer cases and controls tested in three laboratories: a, 1 versus 2 ($r = 0.72$); b, 1 versus 3 ($r = 0.80$); c, 3 versus 2 ($r = 0.61$). The lines portray the linear regression estimate (center line) and the 95% CI around the regression line.

however, there was no preset cutoff. Instead, in laboratory 2, we used receiver-operating characteristic (ROC) analysis in the first part of the study (vulvar cancer cases and controls) to determine a serological cutoff value (1, 20) and then applied this cutoff to the second part of the study (blood donors). Standard software packages, including SAS, SYSTAT/SYGRAPH, PEPI Version 2, and EPI-INFO, were used to conduct all analyses, except for ROC, which was conducted by using ROCLAB (1).

RESULTS

Vulvar cancer patients and controls. Serum samples from 265 of 268 women enrolled were available for testing in both laboratories 1 and 2. The subjects' mean, as well as median, age was 53 years. Caucasian (92%) was the most common race, followed by African American (8%), and there were two Hispanic subjects. Laboratory 3 only tested the first 112 sequential serum samples. However, the demographic and case-control distributions of this subset of subjects were essentially the same as for the study as a whole (data not shown).

The ELISA OD values in the three laboratories are shown in Fig. 1a to c (plotted one laboratory against another over fixed axes to maintain comparability). The geometric mean (interquartile range; a measure of the spread of the data) OD was 0.397 (0.361) in laboratory 1, compared to 0.506 (0.682) in laboratory 2 and 0.180 (0.327) in laboratory 3. Laboratory 1 tested each serum sample three times in separate runs. The triplicates showed good intralaboratory (day-to-day) variability ($\rho \geq 0.88$). Laboratory 3 tested specimens in duplicate but on a single plate. These duplicates showed good intra-assay variability ($\rho = 0.97$).

Interlaboratory agreement was then measured. Linear correlations between each pair of laboratories suggested good agreement. The ρ values were 0.72 (laboratory 1 versus laboratory 2), 0.80 (laboratory 1 versus laboratory 3), and 0.61 (laboratory 2 versus laboratory 3). All P values were <0.0001. Categorical results were determined in laboratories 1 and 3. Laboratory 3 found that 30% of the 112 specimens tested were positive, and seroprevalence was 37% in laboratory 1 in these same samples (Fig. 2). Good interlaboratory agreement between these two laboratories' positive and negative results, in accord with their linear correlation (above), was also found by κ. The κ for laboratory 1 versus laboratory 3 was 0.74 (95% confidence interval [CI], 0.61 to 0.87). Agreement is generally considered excellent for κ values greater than 0.75, fair to good for values between 0.4 and 0.75, and poor for values below 0.4 (3).

A serological OD cutoff value for laboratory 2 was determined a posteriori based on ROC analysis using case-control status and seropositivity in laboratory 1 as separate outcomes. The area under the ROC curve, a measure of the ability of the laboratory 2 assay to discriminate positive from negative results in each outcome, was 0.77 for case-control status and 0.90 for seropositivity in laboratory 1. The latter result is also interpreted to mean that the operating characteristics of the laboratory 1 and 2 assays are similar for these specimens (1, 20).

The serological cutoff in laboratory 2 was set at the lowest OD value that resulted in greater than 90% specificity for both of the outcomes measured. Based on this cutoff, seroprevalence was 31% in the specimens also tested by laboratories 1 and 3. Overall, the three laboratories agreed on 91 (81%; 26

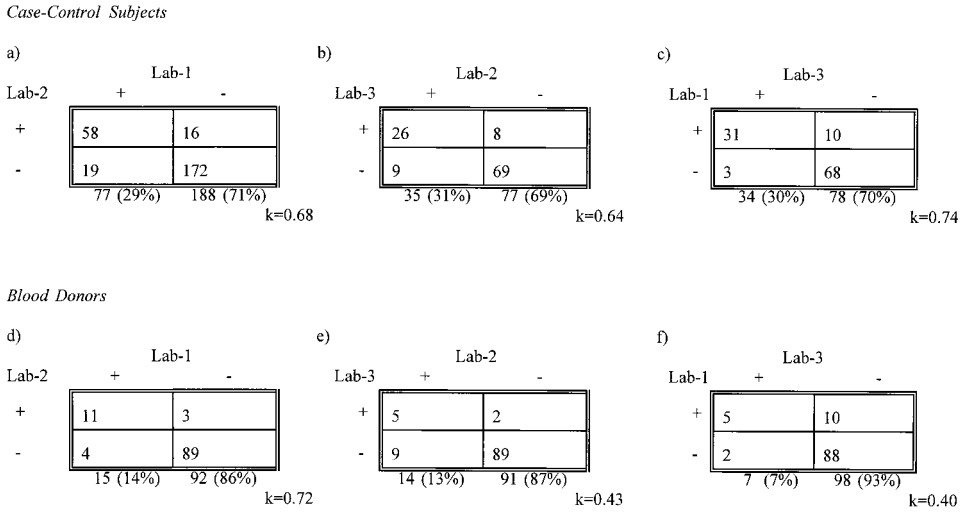


FIG. 2. Positive and negative HPV-16 VLP ELISA findings in three laboratories.

positive and 65 negative) of the 112 specimens. Kappa coefficients showed good agreement between laboratories 2 and 1 ($\kappa = 0.68$; 95% CI, 0.58 to 0.81), between laboratories 2 and 3 ($\kappa = 0.64$; 95% CI, 0.49 to 0.80), and for all three laboratories globally ($\kappa = 0.72$; 95% CI, 0.61 to 0.83). These κ and prevalence results must be interpreted cautiously, however, since it is somewhat circular to determine seroprevalence in laboratory 2, having used this same data set to define the serological cutoff in that laboratory.

Blood donors. All 107 consecutive blood donors were tested, except for two specimens not received in good condition by laboratory 3. Demographic information was available from 96 of these subjects. The mean (median) age of the subjects was 39 years (38 years). Male donors were more common (66%) than females, and Caucasians (59%) were more common than African Americans (40%) or Hispanics (1%).

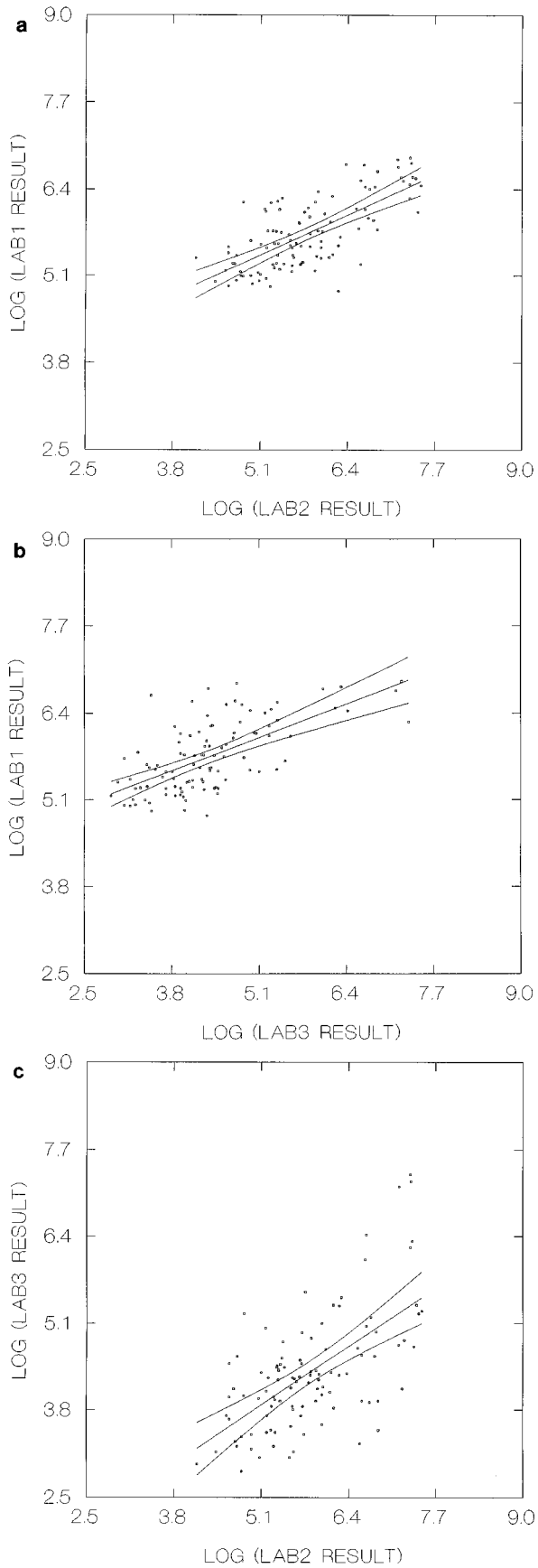
The ELISA OD values in the three laboratories are shown in Fig. 3, plotted one laboratory versus another. The geometric mean (interquartile range) OD values were 0.347 (0.269) in laboratory 1, 0.332 (0.350) in laboratory 2, and 0.077 (0.066) in laboratory 3. Therefore, the average intensity (geometric mean) of serologic responses was less and the spread (i.e., interquartile range) of the data was more limited than for case-control subjects, consistent with expectations for these presumably healthy subjects. Furthermore, since correlation coefficients are sensitive to data spread, it was also expected that the interlaboratory correlations would be somewhat lower for blood donors than for vulvar cancer case-control study subjects. Nonetheless, the Pearson correlation coefficients remained strong; the ρ values were 0.72 (laboratory 1 versus laboratory 2), 0.68 (laboratory 1 versus laboratory 3), and 0.62 (laboratory 2 versus laboratory 3). The intra-assay correlation in laboratory 3 was 0.93. All P values were <0.0001.

Laboratory 1 determined that 12% of the blood donors were seropositive, compared to 14% in laboratory 2 and 7% in laboratory 3. Overall, the three laboratories agreed on 90 (86%; 4 positive and 86 negative) of 105 specimens, similar to the interlaboratory concordance found in case-control study subjects. Kappa statistics for the categorical results also showed good agreement between laboratories 1 and 2 ($\kappa = 0.72$; 95% CI, 0.53 to 0.92). However, κ showed only fair agreement between laboratories 3 and 1 ($\kappa = 0.42$; 95% CI, 0.14 to 0.71), as well as between laboratories 3 and 2 ($\kappa = 0.40$; 95% CI, 0.12 to 0.68), inconsistent with their strong linear correlations. The global kappa coefficient for all three laboratories showed fair-to-good agreement ($\kappa = 0.53$; 95% CI, 0.42 to 0.64).

We previously noted (see Materials and Methods) that agreement is underestimated by κ in skewed distributions, and in this connection, laboratory 3 had the most uneven distribution of positive and negative results. However, we also wanted to know if interlaboratory agreement in categorical data among blood donors was affected by the strength of the antibody reactions (i.e., a paucity of clearly positive results). To this end, we divided serologic responses a posteriori by their level of reactivity. We defined strongly positive sera in each laboratory as those with an ELISA OD value above the median for seropositive case-control study subjects. Similarly, we defined clearly negative sera as any with a value below the median for seronegative case-control study subjects. Relatively few blood donor serum specimens were strongly positive; 3 to 5% of all blood donors, compared to 13 to 15% of all case-control study subjects, were strongly positive in each laboratory. When analysis of blood donor samples was limited to strongly positive and clearly negative sera, the results from laboratory 3 agreed with findings from laboratory 1 for 93% of the samples and with the results from laboratory 2 for in 95% of the samples. In contrast, when we studied blood donor sera that did not give strongly positive or clearly negative results in laboratory 3, the overall agreement was 77% with laboratory 1 and 64% with laboratory 2. Among both case-control study subjects and blood donors, 55 of the 56 serum samples that gave either strongly positive or clearly negative results in all three laboratories had concordant results in all three laboratories.

DISCUSSION

This is the first study to assess interlaboratory agreement in HPV-16 VLP ELISA results, despite the increasingly frequent use of these ELISAs in seroepidemiologic studies (2, 11, 12, 14, 15). Interlaboratory agreement is a concern in any seroassay that has not undergone extensive comparative field testing. However, it is a particular concern in measuring antibodies to



HPV. HPV infections are confined to the epithelium and do not have a viremic phase (8). As a result, serum antibodies to HPV are present only at low levels and positive specimens often may be close to the margin of detectability. In this study, we compared HPV-16 VLP ELISA results in three laboratories that tested a heterogeneous mixture of specimens from two separate populations—subjects from a vulvar cancer case-control study and healthy male and female blood donors.

Our results suggested that there is good intralaboratory (day-to-day) agreement and moderate-to-good interlaboratory agreement in the HPV-16 VLP ELISA results. Specifically, strong intralaboratory agreement was suggested by the high Pearson correlations ($\rho \geq 0.87$) in triplicate samples tested in separate runs in laboratory 1. Similar data were not available in the other testing laboratories, but the strong intra-assay correlation ($\rho \geq 0.93$) in laboratory 3 adds further evidence that each laboratory can replicate its own results. Likewise, good interlaboratory agreement was suggested by strong Pearson correlations and ROC results. Interlaboratory Pearson correlation values ranged between 0.61 and 0.80 for both case-control study subjects and healthy blood donors, indicating that ELISA OD levels in the three laboratories were linearly related regardless of the population. Consistent with these results, ROC analysis demonstrated that at least two of the assays had similar operating characteristics (the area under the ROC curve was 0.9) (20). As a whole, the direct comparison of OD levels in the three laboratories supports the conclusion that each laboratory measured the same humoral immune response and that their assays performed similarly.

In separate analyses, interlaboratory agreement was also measured according to the categorical interpretation of ELISA values in each laboratory. These results gave a generally encouraging, albeit slightly more mixed, picture of interlaboratory agreement. That is, in accord with linear regression, among case-control study subjects the κ values were universally strong (>0.6) across the three laboratories, and for blood donors, the κ values showed strong interlaboratory agreement between laboratories 1 and 2 (3). However, κ suggested only fair agreement for blood donors between laboratory 3 and the other two laboratories. Thus, for blood donors in laboratory 3 there was a disparity between linear correlation and κ. We have previously noted that part of this disparity might be artifactual (18), in that κ underestimates agreement in skewed distributions, such as seroprevalence among blood donors, which was low in all three laboratories and was incrementally the lowest in laboratory 3. More interesting, however, was the observation that a low frequency of strong positive results may have added to the variability of the categorical data on healthy blood donors. These findings suggest that by excluding intermediate values (i.e., defining a range of indeterminate OD values), interlaboratory agreement can be improved. More work to define an optimal range of indeterminate values is clearly needed, but that will best be accomplished in studies specifically designed for that purpose.

Another way to improve interlaboratory agreement is by sharing positive and negative control sera. To this end, the Viral Epidemiology Branch of the National Cancer Institute will maintain a panel of highly characterized serum specimens which will be made available to the research community. Optimal interlaboratory agreement, however, will probably only

FIG. 3. Comparison of HPV-16 VLP ELISA OD values for healthy blood donors tested in three laboratories: a, 1 versus 2 ($r = 0.72$); b, 1 versus 3 ($r = 0.68$); c, 3 versus 2 ($r = 0.62$). The three lines portray the linear regression estimate (center line) and the 95% CI around the regression line.

be obtained if a commercial ELISA kit is developed which will help to minimize variations in laboratory procedures and reagents and set a serological cutoff.

In summary, the HPV-16 VLP ELISA has demonstrated its value as a seroepidemiologic tool in many studies (2, 11, 12, 14, 15). Intralaboratory agreement is good, and interlaboratory agreement is moderate to good. The assay, therefore, is currently best used to make comparisons between one population and another (e.g., cases and controls) within a single laboratory. However, improvement in the generalizability of results from one laboratory to another (e.g., in measuring population seroprevalence) can be made by treating marginal values as indeterminate and by sharing reference positive and negative serum specimens throughout the research community.

REFERENCES

1. **DeLeo, J. M.** 1993. Receiver operating characteristics laboratory (ROCLAB): software for developing decision making strategies that account for uncertainty, p. 218–325. *In* Proceedings of the Second International Symposium on Uncertainty in Modeling and Analysis. IEEE Computer Society, Los Alamos, Calif.
2. **Dillner, J., F. Wiklund, P. Lenner, C. Eklund, V. Frederiksson-Shanazarian, J. T. Schiller, M. Hibma, G. Hallmans, and U. Stendahl.** 1995. Antibodies against linear and conformational epitopes of human papillomavirus type 16 that independently associate with incident cervical cancer. Int. J. Cancer **60:**377–382.
3. **Fleiss, J. L.** 1981. Statistical methods for rates and proportions, p. 212–236. John Wiley & Sons, Inc., New York, N.Y.
4. **Galloway, D. A.** 1995. Papillomavirus capsids: a new approach to identify serological markers of HPV infection. JNCI **86:**474–475.
5. **Heino, P., C. Eklund, V. Fredriksson-Shanazarian, S. Goldman, J. T. Schiller, and J. Dillner.** 1995. Association of serum immunoglobulin G antibodies against human papillomavirus type 16 capsids with anal epidermoid carcinoma. JNCI **87:**437–440.
6. **Hildesheim, A., C. L. Han, L. A. Brinton, P. C. Nasca, R. M. Richart, C. L. Trimble, R. J. Kurman, R. B. Jones, R. L. Ashley, R. G. Ziegler, K. Mullin, and J. T. Schiller.** Human papillomavirus type 16 and vulvar cancer risk: results from a seroepidemiological case-control study. Cancer Causes Control, in press.
7. **IARC Working Group.** 1996. Monographs on the evaluation of carcinogenic risks to humans. Vol. 64: human papillomaviruses. IARC, Lyon, France.
8. **I-Kuei Lin, L.** 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics **45:**255–268.
9. **Kirnbauer, R., F. Booy, N. Cheng, D. R. Lowy, and J. T. Schiller.** 1992. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proc. Natl. Acad. Sci. USA **89:**12180–12184.
10. **Kirnbauer, R., J. Taub, H. Greenstone, R. Roden, M. Durst, L. Gissmann, D. R. Lowy, and J. T. Schiller.** 1993. Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. J. Virol. **67:**6929–6936.
11. **Kirnbauer, R., N. L. Hubbert, C. M. Wheeler, T. M. Becker, D. R. Lowy, and J. T. Schiller.** 1994. A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. JNCI **86:**494–499.
12. **Nonnenmacher, B., N. L. Hubbert, R. Kirnbauer, K. V. Shah, N. Munoz, F. X. Bosch, S. de Sanjose, R. Viscidi, D. R. Lowy, and J. T. Schiller.** 1995. Serologic response to human papillomavirus type 16 (HPV-16) virus-like particles in HPV-16 DNA-positive invasive cervical cancer and cervical intraepithelial neoplasia grade III patients and controls from Colombia and Spain. J. Infect. Dis. **172:**19–24.
13. **Palefsky, J. M.** 1995. Serologic detection of human papillomavirus-related anogenital disease: new opportunities and challenges. JNCI **87:**401–402.
14. **Strickler, H. D., M. H. Schiffman, K. V. Shah, C. S. Rabkin, J. T. Schiller, S. Wacholder, B. Clayman, and R. P. Viscidi.** Submitted for publication.
15. **Strickler, H. D., M. H. Schiffman, C. Eklund, A. G. Glass, D. R. Scott, M. E. Sherman, S. Wacholder, R. J. Kurman, M. M. Manos, J. Schiller, and J. Dillner.** 1997. Evidence for at least two distinct groups of humoral immune reactions to papillomavirus antigens in women with SIL. Cancer Epidemiol. Biomark. Prev. **6:**183–199.
16. **Trimble, C. L., A. Hildesheim, L. A. Brinton, K. V. Shah, and R. J. Kurman.** 1996. Heterogeneous etiology of squamous carcinoma of the vulva. Obstet. Gynecol. **87:**59–64.
17. **Viscidi, R. P., K. L. Kotloff, B. Clayman, K. Russ, S. Shapiro, and K. V. Shah.** 1997. Prevalence of antibodies to human papillomavirus (HPV) type 16 virus-like particles in relation to cervical HPV infection among college women. Clin. Diagn. Lab. Immunol. **4:**122–126.
18. **Whitehurst, G. J.** 1984. Interrater agreement for journal manuscript reviews. Am. Psychol. **39:**22–28.
19. **Wideroff, L., M. H. Schiffman, B. Nonnenmacher, N. Hubbert, R. Kirnbauer, C. E. Greer, D. Lowy, A. T. Lorincz, M. M. Manos, A. G. Glass, D. R. Scott, M. E. Sherman, R. J. Kurman, J. Buckland, R. E. Tarone, and J. T. Schiller.** 1995. Evaluation of seroreactivity to HPV 16 virus-like particles in an incident case-control study of cervical neoplasia. J. Infect. Dis. **172:**1425–1430.
20. **Zewig, M. H., and G. Campbell.** 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. **39:**561–577.